**Regulation of bovine papillomavirus type 1 gene expression by RNA processing**

**Rong Jia and Zhi-Ming Zheng**

*HIV and AIDS Malignancy Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD*

**TABLE OF CONTENTS**

## 1. ABSTRACT

Bovine papillomavirus type 1 (BPV-1) has served as a prototype for studying the molecular biology and pathogenesis of papillomaviruses. The expression of BPV-1 early and late genes is highly regulated at both transcription and post-transcriptional levels and strictly tied to the differentiation of keratinocytes. BPV-1 infects keratinocytes in the basal layer of the skin and replicates in the nucleus of infected cells in a differentiation-dependent manner. Although viral early genes begin to be expressed from the infected, undifferentiated basal cells, viral late genes are not expressed until the infected cells enter the terminal differentiation stage. Both BPV-1 early and late transcripts are intron-containing bicistronic or polycistronic RNAs, bearing more than one open reading frame and are polyadenylated at either an early or late poly (A) site. Nuclear RNA processing of these transcripts by RNA splicing and poly (A) site selection has been extensively analyzed in the past decade and various viral cis-elements and cellular factors involved in regulation of viral RNA processing were discovered, leading to our better understanding of the gene expression and biology of human papillomaviruses.

## 2 INTRODUCTION

Bovine papillomavirus type 1 (BPV-1) is a small DNA tumor virus that causes benign fibropapillomas in cattle. BPV-1 has an 8-kb double-stranded circular DNA genome (1) which can be separated into three distinguishable regions: a long control region (LCR) or noncoding region (NCR), an early or transforming region, and a late region encoding structural capsid proteins. These three regions are divided by two polyadenylation (pA) sites: early pA (AE) and late pA (AL). The early region encodes six early open reading frames (ORFs): E1, E2, E5, E6, E7, and E8, and a late ORF, E4 (2-5); the late region encodes two ORFs: L1 and L2. All ORFs are located on one strand of the double-stranded genome. The LCR contains a DNA replication origin (6,7), four promoters (P7185, $P_L$, $P_{7940}$, and $P_{89}$,), and 12 enhancer elements (8-12) which are activated by the full-length E2 protein. Transcription of BPV-1 genes is very complex. The early region transcripts of BPV-1 are transcribed from at least six promoters: $P_{7185}$, $P_{7940}$, $P_{89}$, $P_{890}$, $P_{2443}$, and $P_{3080}$, but all of these early transcripts are polyadenylated at the same $A_E$ site (Figure 1).
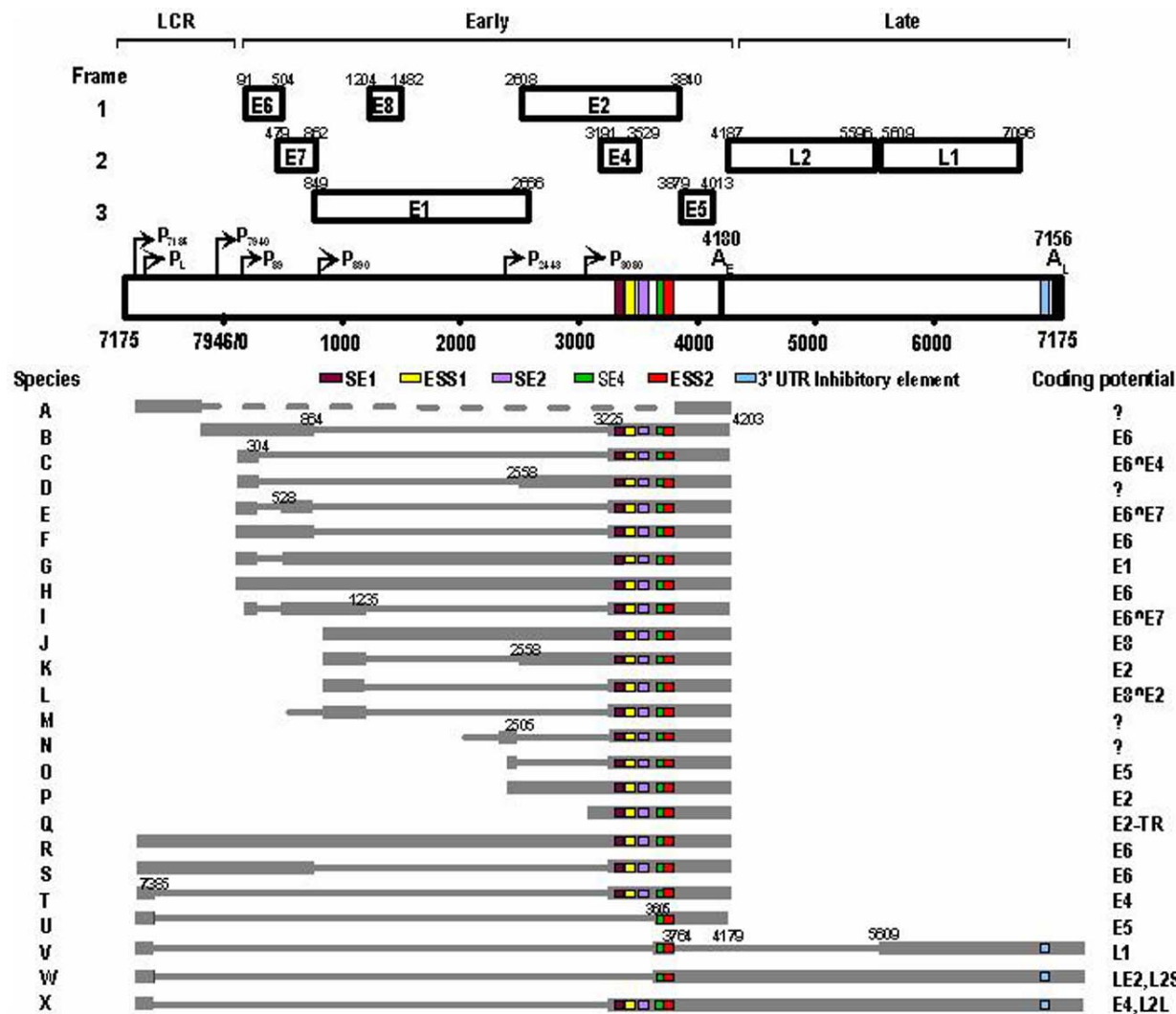
**Figure 1.** Genome structure and transcription map of BPV-1. Showing on the top is the genomic organization of BPV-1. The bracket line in the middle of the figure represents a linear form of the virus genome, showing seven designated promoters (arrows), with the nucleotide position of the RNA start site for a given promoter, and two polyadenylation sites ($A_E$ and $A_L$). The open reading frames (ORFs) (open boxes) are diagramed above the bracket and the numbers above each ORF are the nucleotide positions of the first nucleotide of the start codon and last nucleotide of the stop codon in the BPV-1 genome. The reported RNA species derived from individual promoter and their splicing directions are showed below the bracket. Heavy lines and thin lines represent exons and introns, respectively, in each species of the RNA, with splice site positions being numbered by the nucleotide positions in the virus genome. Each cis-element is depicted at its relative position of each RNA species and labeled with its corresponding color. Figure is modified with permission from reference (83).

The $P_{89}$ promoter is a classical promoter, with a TATA box about 30 nucleotides (nt) upstream of the E6 initiation site and a CAAT-box at position -83 (13). The $P_{89}$ is one of the most active promoters in BPV-1 transformed cells and can be transactivated by E2 *in vivo* (9,14,15). However, an upstream E2 responsive enhancer (E2RE) is not required for efficient transcription *in vitro* from the $P_{89}$ (16). Three transcription factor Sp1 sites are located at nt 7800, 7833 and 7854 upstream of the $P_{89}$ start site. These Spl sites are critical for basal and E2-transactivated transcription from the $P_{89}$ promoter (17). The $P_{89}$ transcripts are responsible for the expression of transforming E6 (14) and possibly for E7 (18,19). A spliced isoform (nt 304/528)

of the $P_{89}$ transcripts presumably results in the production of an E6^E7 fusion (14,15), but this has never been confirmed.

Transcripts from $P_{2443}$ and $P_{3080}$ are the most abundant BPV-1 transcripts in the fibromatous portion of the fibropapilloma (13). $P_{2443}$ does not have a classical TATA box, but does have the sequence TAATATT located about 30 nts upstream of the transcription start site. Two transcripts are expressed from $P_{2443}$. An unspliced message encodes a full-length E2 protein and a spliced form (nt 2505/3225) encodes the E5 protein. The spliced E5 form accounts for 90% of transcripts (14,15). $P_{2443}$ is also

transactivated by E2 through the E2 responsive enhancer elements in LCR (20). A Sp1 binding site upstream of the $P_{2443}$ TATA-like box is critical for the basal level of transcription from this promoter (21). $P_{3080}$ is located in the middle of the E2 ORF and does not have any recognizable TATA box (13,22). Like $P_{7940}$, $P_{89}$, and $P_{2443}$, activity of $P_{3080}$ is also regulated by the E2 protein. $P_{3080}$ transcripts encode a transcriptional repressor, E2-TR, that contains 162 amino acid (aa) residues from the C-terminal half of the E2 ORF and binds the same E2 responsive enhancer elements as the full-length E2 protein, but inhibits E2-dependent transactivation (3,4). The $P_{890}$ transcripts also encode another transactivation repressor E8^E2. The E8^E2 fusion protein has 11 aa N-terminal residues from the E8 ORF and 207 aa C-terminal residues from the E2 ORF. This fusion is created from a spliced message transcribed from the $P_{890}$ promoter by RNA splicing from nt 1235 to 3225 (4,5,23).

Two additional promoters, $P_{7185}$ and $P_{7940}$ in the LCR are also active in both bovine warts and BPV-1 transformed mouse C127 cells (ID13) (13). $P_{7185}$ is the weakest BPV-1 promoter *in vivo* and does not have a classical TATA box, although a potential TATA box is located 30 nts upstream from the RNA start site, which overlaps with the AATAAA polyadenylation signal for the late region RNAs (13,16). $P_{7185}$ has rather strong activity in an *in vitro* HeLa whole cell extract transcription system and responses to cycloheximide *in vivo* (13,16). In contrast to $P_{7185}$, $P_{7940}$ has a TATA box and a CCAAT box upstream of the transcription start site. $P_{7940}$ contains an E2 responsive enhancer and is responsive to E2 transactivation (9,24). The level of transcripts expressed from $P_{7940}$ is approximately one-third the amount from $P_{89}$ (13). Deletion of the sequence between the TATA box and the transcription start site of the $P_{7940}$ promoter reduces the expression of E1 (25).

BPV-1 late region mRNAs are transcribed from a late promoter $P_L$. In a fibropapilloma, the majority of transcripts from $P_L$ have 5' termini near nt 7250 and utilize a common splice donor site at nt 7385. The $P_L$ is not utilized in the BPV-1-transformed C127 cell or in fibromatous portion of a fibropapilloma, but the activity of the $P_L$ promoter dramatically increases in the granular layer of a fibropapilloma (26) where viral DNA replication takes place. Upstream of this mRNA start site is a tandemly repeated sequence element homologous to the SV40 late promoter sequence. (13,27,28).

## 3. BPV-1 GENE EXPRESSION IS REGULATED BY RNA SPLICING

After promoter induction and initiation of viral transcription, removal of introns from a nascent pre-mRNA transcript by RNA splicing is a major step toward maturation of mRNA. Thus, regulation of RNA splicing at the posttranscriptional level plays an important role during the life cycle of BPV-1 and other papillomaviruses. BPV-1 transcripts are either bicistronic or polycistronic and contain only one type of intron: the U2-type GU-AG or GC-AG intron (Figure 2), which is the most common intron

in all eukaryotic transcripts. The U12-type AU-AC intron is rare in eukaryotic transcripts and has not been identified in any of the papillomavirus transcripts to date. The bicistronic or polycistronic features of BPV-1 transcripts, composed of several ORFs in each transcript (Figure 1), are common in RNA transcripts of all other papillomaviruses. In addition, alternative RNA splicing, which disrupts or removes an ORF from a given transcript to direct a particular ORF usage for protein translation (29) is an evolutionarily conserved function in the gene expression of all papillomaviruses.

### 3.1. Splicing of viral early transcripts

Transcripts from the $P_{89}$ promoter have two introns and three exons. The RNAs spliced from nt 304 to 528 cannot encode the E6 or E7 protein, but the transcripts with this intron retention can only encode E6 protein. The question remains whether this transcript encodes E7 protein. While it is presumed that E1, E6, and E7 proteins are made in basal cells from transcripts originating at promoter $P_{89}$, this has yet to be unproven. In BPV-1 transformed mouse C127 cells, a transcript with a 5' end at nt 429 was hypothesized to direct the synthesis of E7 protein, suggesting the presence of a separate putative promoter for E7 transcription (15). However, this hypothesis was not validated in a separate study (22). The second intron between nt 865 and 3224 of $P_{89}$ transcripts is spliced usually, but can be retained in a small portion of the transcripts. Splicing from nt 304 to 3225 in $P_{89}$ transcripts will delete the exon from nt 528 to 864 and create a transcript which has potential to encode a fusion protein E6^E4 using an AUG initiation codon at nt 91-93 and a UGA termination codon at nt 3527-3529 (15). Another $P_{89}$ transcript spliced from nt 304 to 2558 may encode a full-length E2 protein (14,30).

The E1 ORF is the largest in the BPV-1 genome and its expression is essential for viral DNA replication initiation and elongation. E1 proteins are well-conserved among bovine and human papillomaviruses. Transcripts containing the E1 ORF region without splicing are very rare, but are detectable in cycloheximide-treated, BPV-1-infected bovine cells by a probe from a central portion of the E1 region (31). The E1 protein (68 kDa) is also detectable in BPV-1 transformed mouse C127 cells (32). The intron 2 in the E1 ORF has an alternative 5' splice site (5' ss), nt 1235 5' ss. Splicing of the transcript from the nt 1235 5' ss to nt 3225 3' splice site (3' ss) produces a transcript that encodes a 23-kDa E1M protein, which contains 129 amino acid residues from E1 and 13 amino acid residues from downstream sequences outside of E1 ORF (5,32). However, the biological function of E1M protein to papillomavirus replication remains uncertain (33).

Transcripts expressed from the $P_{890}$ promoter may splice from nt 1235 5' ss to two alternative 3' ss either at nt 2558 or at nt 3225. Transcripts using nt 3225 3' ss encode a transcriptional repressor E8^E2 (5). Transcripts using nt 2558 3' ss have the potential to express a full-length E2 protein. These transcripts can be found in mouse cells transfected by BPV-1 DNA and the level of
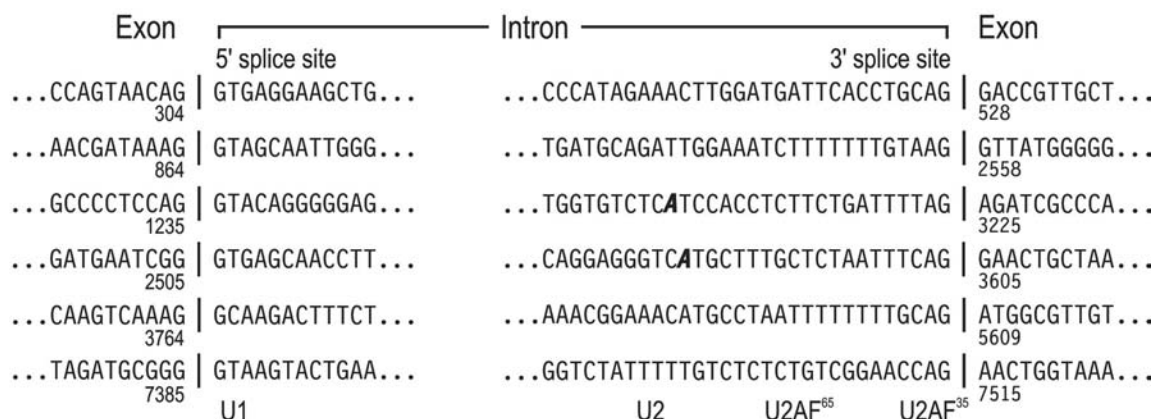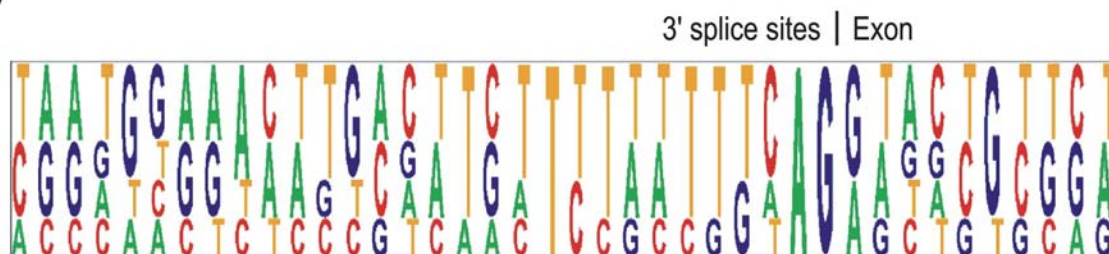
**Figure 2.** Splice sites of BPV-1 RNA transcripts. (A) Sequences of 5' splice sites and 3' splice sites in BPV-1 pre-mRNAs. T = U in RNA. Exon and intron juctions are separated by vertical lines. Numbers below each sequence are the nucleotide positions in the virus genome and indicate either the last nucleotide of the exon immediately upstream of an intron 5' splice site or the first nucleotide of the exon immediately downstream of an intron 3' splice site. A mapped branch point nucleotide in the nt 3225 3' splice site (45) and the nt 3605 3' splice site (49) is bolded in Italic. Showing below the sequences are relative positions for U1, U2, U2AF[65], and U2AF[35] binding. (B and C) Pictograms of BPV-1 5' splice sites (B) and 3' splice sites (C) showing the relative frequencies of the bases at each position (http://genes.mit.edu/pictogram.html). T = U in RNA. The nt 7175 3' splice site in (A) was not included for the pictogram analysis because this cryptic 3' splice site is activated only when the 3225 3' splice site is mutated (49).

expression is the same as in the unspliced transcripts expressed from the $P_{2443}$, which also encode the full-length E2 transactivator (30). Transcripts from the $P_{2443}$ promoter spliced from nt 2550 to 3225 encode the E5 oncoprotein which is required for cell transformation.

### 3.2. Splicing of viral late transcripts

The BPV-1 life cycle is tightly linked to the differentiation state of their host keratinocytes (26,34). The structural proteins, L1 and L2, are only expressed in cells of the granular or most differentiated layer of the epithelium. The $P_L$ promoter is activated in both the upper spinous and granular cell layers of the bovine fibropapilloma (26). Transcripts from the late promoter $P_L$ start at nt 7250 and share the common immediate downstream 5' ss at nt 7385, but utilize two alternative 3' ss, nt 3225 3' ss or nt 3605 3' ss, for their expression (34). Splicing from nt 7385 to 3225 and polyadenylation at the

$A_L$ site form a large mRNA to encode L2 protein. Splicing from nt 7385 to 3605 activates the removal of another intron downstream from nt 3765 to 5608 (26). This alternative splicing produces a smaller transcript that encodes L1, a major capsid protein. However, a small portion of the late transcripts that use the nt 3605 3' ss for splicing does not remove the downstream intron and consequently produce an RNA species L2-S. This transcript is smaller in size than the L2 RNA (also called L2-L) which is spliced at the nt 3225 3' ss. Northern analysis of bovine fibropapilloma tissues has shown that both L1 and L2 mRNAs are abundantly expressed transcripts (26).

ORF E4 completely overlaps with ORF E2 in a different open reading frame. The function of BPV-1 E4 is unknown. E4 does not promote mouse C127 cell transformed focus formation and has no transcriptional activities (35). Transcripts from the $P_L$ or $P_{7185}$ promoters spliced from nt 7385 to 3225 and polyadenylated at the $A_E$ have the potential to encode E4 (13); those spliced at nt 3605 have the potential to encode E5 if polyadenylated at the $A_E$ site. Several other splice sites, such as nt 1613 5' ss and nt 1940 5' ss, and nt 1866 3' ss, nt 1032 3'ss, and nt 1024 3' ss in the E1 region are rarely utilized, but have been identified only in ID13 cells (5).

### 3.3. Splicing elements in BPV-1 transcripts

A splicing element is any element in an RNA transcript that is required for RNA splicing. Prior to 1993, these splicing elements were usually those elements in the RNA introns that are necessary for spliceosome-mediated intron recognition and splicing. These include an intron 5' end for U1 binding (Figure 2A and 2B) and several elements at RNA intron 3' end important for the identification and accurate exclusion of an intron from a pre-mRNA. The intron 3' end has three cis elements: a branch point (BP), followed by a polypyrimidine tract (PPT), and a terminal AG dinucleotide. During spliceosome assembly, $U2AF^{65}$ and $U2AF^{35}$ bind to the polypyrimidine tract and terminal AG dinucleotide, respectively, and facilitate binding of U2 to the branch point sequence via its RNA base-pairing (Figure 2A) (36-42). However, identification of an exon recognition sequence which enhances the splicing of a GU-AG intron in 1993 (43) was surprising and elicited immediately a search for exonic elements that stimulate efficient RNA splicing. This search led to the discovery of a series of exonic splicing enhancers (ESEs) and exonic splicing suppressors or silencers (ESSs) that regulate alternative splicing of BPV-1 late transcripts (44).

As described above, BPV-1 L1 and L2 are transcribed from the same late promoter, $P_L$, as a bicistronic transcript that contains three exons and two introns. Intron 1 of the late transcripts is a typical GU-AG intron and bears two alternative 3' ss, at nt 3225 and nt 3605, whereas the intron 2 is a GC-AG intron (Figure 2A) and has L2 ORF within it. In bovine warts, selection of the nt 3225 3' ss allows retention of intron 2 and L2 expression. In contrast, selection of the nt 3605 3' ss can activate splicing of the intron 2 leading to L1 expression. Notably, the nt 3225 3' ss is also a common 3' ss for intron 2 splicing of all early viral

transcripts (Figure 1). Characterization of BPV-1 nt 3225 and 3605 3' ss have shown that both nt 3225 and 3605 3' ss are functionally suboptimal and have a nonconsensus BP sequence and a weak PPT interspersed with purines (Figure 2A and 2C). Thus, the selection of these suboptimal 3' ss is regulated by ESEs and ESSs (45). Five cis elements are involved in the regulation of the selection of the two alternative 3' ss, including three ESEs (SE1, SE2, and SE4) and two ESSs (ESS1 and ESS2) (Figure 3). SE1, ESS1, and SE2 are located between the nt 3225 3' ss and nt 3605 3' ss (Figure 3A). SE1 and SE2 are purine-rich ESEs and synergistically promote the selection of nt 3225 over the suppression by the ESS1, which contains a suppressive core motif GGCUCCCCC (45-48). With respect to the nt 3605 3'ss, SE1 and SE2 are also intronic elements and may have the potential to regulate usage of the nt 3605 3' ss. SE4 and ESS2 are located between nt 3605 3' ss and nt 3764 5' ss. SE4 is an AC-rich element and promotes selection of the nt 3605 3' ss over the suppression by ESS2. ESS2 contains a UGGU core suppressor motif, which may base pair with the functional CACCACCAC motif of SE4. Two copies of UGGU are more inhibitory for *in vitro* RNA splicing (49). Although SE4 may function as an ESE in the context of different 3' ss, ESS2 functions as an ESS only in a splice site-specific manner (49). This feature of the ESS2 is in contrast to the ESS1 which functions in a less splice site-specific manner and inhibits not only 3225 3' ss in BPV-1 late pre-mRNA, but also suboptimal 3' ss in human immunodeficiency virus tat-rev pre-mRNA, *Drosophilia* dsx pre-mRNA, and Rous sarcoma virus src pre-mRNA. Optimization or strength of 3225 3' ss counteracts the repression function of ESS1 (48).

Whether a splicing element is required to regulate alternative splicing of early viral transcripts remains unknown. As the nt 3225 3' ss is also a major 3' ss for many early viral transcripts, its utilization in splicing of viral early transcripts, by analogy, might be also associated with the activities of SE1, ESS1, and SE2, three exonic cis-elements in regulation of alternative RNA splicing of viral late transcripts (Figure 1). This may be also true for the expression of viral E2. The E2 protein plays an important role in viral DNA replication and transactivation of viral genes. As we discussed previously, two types of RNA transcripts, $P_{890}$ and $P_{2443}$, have the potential to encode a full-length E2 that functions as a transcriptional transactivator (E2-TA). The $P_{890}$ transcript contains an intron with a 5' ss at nt 1235 and two alternative 3' ss; one 3' ss position at nt 2558 and another at nt 3225 (Figure 4). Thus, the $P_{890}$ transcripts can splice from the nt 1235 5' ss to either nt 2558 3' ss for the expression of a full-length E2-TA or nt 3225 3' ss for the expression of an E8^E2 fusion that functions as a transcriptional repressor. The $P_{2443}$ transcript has an intron with a 5' ss at nt 2505 and a 3' ss at nt 3225. Retention of this intron in the transcript leads to the expression of a full-length E2-TA (Figure 4), but splicing of this intron can result in the production of E5 (Figure 1). Thus, it is conceivable that a delicate regulation in control of RNA splicing of these two transcripts would be essential for a well balanced expression of E2-TA, E8^E2, and E5.
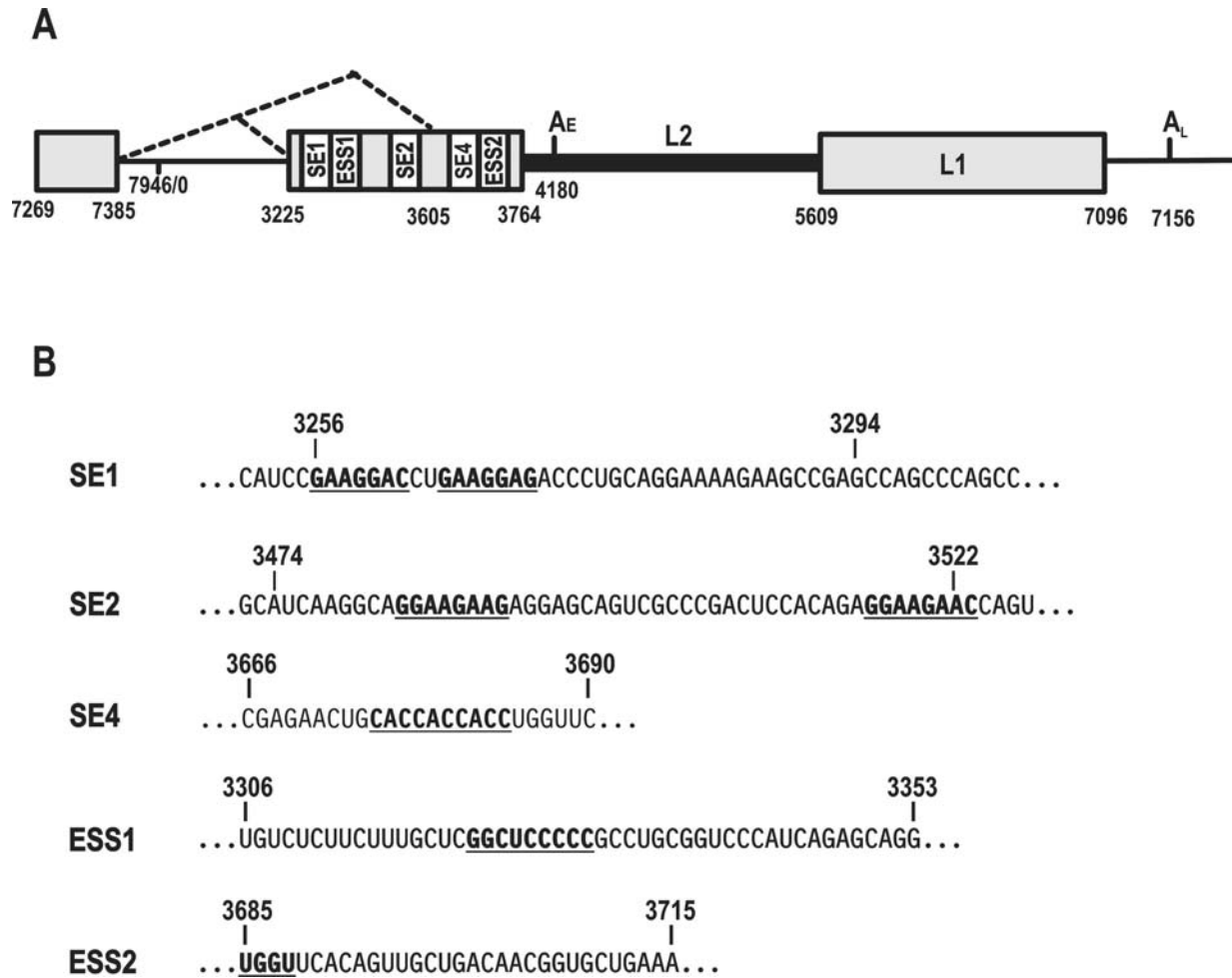
**Figure 3.** Cis elements in regulation of BPV-1 late pre-mRNAs. (A) A BPV-1 late pre-mRNA with five cis elements identified in its exon 2. The expression of BPV-1 late genes is regulated by alternative splicing. Splicing at nt 3225 3' splice site leads to production of a large mRNA to encode L2 protein, but splicing at nt 3605 3' splice site removes the downstream intron from nt 3765 to 5608 (L2 ORF) and produces a smaller transcript for translation of L1. The selection of the two alternative 3' splice sites is determined by three exonic splicing enhancers (SE1, SE2, and SE4) and two exonic splicing suppressors (ESS1 and ESS2). (B) The sequences highlighted by underlines in bold are the functional motifs of individual cis-element. SE1, ESS1, and SE2 regulate the selection of nt 3225 3' splice site and SE4 and ESS2 contribute to the selection of nt 3605 3' splice site.

### 3.4. Cellular splicing factors in BPV-1 RNA splicing

BPV-1 SE1 and SE2 are purine-rich ESEs and each has two ASF/SF2 (alternative splicing factor/splicing factor 2) -binding sites similar to the ASF/SF2 consensus sequence RGAAGAAC (R=A or G): GAAGGAC and GAAGGAG for SE1 and GGAAGAAG and GGAAGAAC for SE2 (Figure 3B) (46). Binding of ASF/SF2 to SE1 and SE2 enhances RNA splicing at the nt 3225 ss. Cells with a high level of ASF/SF2 increase usage of the nt 3225 ss, over the nt 3605 3'ss (50). In fact, ASF/SF2 expression in the upper epithelial layers with highly differentiated keratinocytes is significantly reduced when compared with the levels in basal or superbasal layers with undifferentiated keratinocytes (51). Other SR (serine/arginine-rich) proteins also bind SE1 and SE2. These proteins are SC35, SRp55, and SRp75 (46) (Jia R and Zheng ZM, unpublished data). However, the binding motifs for these proteins in both SE1 and SE2 remain unidentified.

BPV-1 ESS1 is a 48-nt pyrimidine-rich sequence motif (Figure 3B) and has a U-rich 5' region interacting with U2AF and PTB proteins, a central C-rich part binding 35- and 54- to 55-kd SR proteins, and an AG-rich 3' end interacting with ASF/SF2. In general, the PTB binding site overlaps the U2AF[65] binding site in the PPT region of RNA 3' ss. Binding of PTB to this site prevents the binding of U2AF[65] to PPT and negatively regulates splicing of the downstream 3' ss. Although both U2AF[65] and PTB bind the ESS1 5' U-rich region, their binding to this region is not essential for ESS activity. Instead, the most critical region of ESS1 is the central C-rich core GGCUCCCC. Binding of SR proteins to the central core, along with additional SR protein binding to the sequence immediately downstream of the core, is sufficient for partial suppression of spliceosome assembly and splicing of BPV-1 pre-mRNAs (47). It remains unknown what binds ESS2 and executes its negative role in selection of the nt 3605 3' ss.

As an AC-rich ESE, BPV-1 SE4 interacts with SRp20 and YB-1, and regulates alternative RNA splicing of viral late pre-mRNAs. Binding of SRp20 to SE4 suppresses selection of the BPV-1 late-specific nt 3605 3' ss for L1 expression, but favors usage of the nt 3225 3' ss for splicing to express viral L2 and early proteins. Thus, cells with deficient expression of SRp20 have more L1 expression. In terminal differentiated keratinocytes of bovine wart tissues, L1 expression inversely correlates with levels of SRp20 (Jia R and Zheng ZM, Unpublished data). In contrast, binding of YB-1 to SE4 has only a minimal, positive effect on selection of the nt 3605 3' ss.

## 4. REGULATION OF BPV-1 GENE EXPRESSION BY ALTERNATIVE POLYADENYLATION

RNA polyadenylation plays an important role in the control of eukaryotic and viral gene expression. This process involves cleavage of the nascent transcript (52,53) and addition of a poly (A) tail with 150-200 adenylate residues following RNA splicing in the nucleus. Cleavage and polyadenylation are tightly coupled events that are triggered through recognition of three RNA signals by the cellular polyadenylation machinery: a highly conserved AAUAAA hexamer, a cleavage site generally positioned 10-30 nt downstream of AAUAAA, and a G/U- or U-rich element that is ~10-30 nt further downstream of the cleavage site (54). Cellular polyadenylation factors include the cleavage and polyadenylation specificity factor (CPSF) that binds to the AAUAAA hexamer, the cleavage stimulation factor (CstF) that binds to the G/U- or U- rich sequence, the cleavage factors I and II (CFI and CFII) that bind to the cleavage site, poly (A) polymerase (PAP) for poly (A) addition, and the poly (A) binding protein (PABP) that associates with the nascent poly (A) tail to facilitate RNA export and translation (55,56).

The BPV-1 genome contains two major polyadenylation signals: the major early polyadenylation signal $A_E$ at nt 4180 and the major late polyadenylation signal $A_L$ at nt 7156. Recognition of the $A_E$ at nt 4180 promotes the cleavage of all early transcripts at nt 4203 for addition of a poly (A) tail (15) and precludes the expression of viral L1 and L2 genes. This polyadenylation activity requires that all viral early transcripts are terminated before reaching the viral late $A_L$ signal and are in preferential selection of the viral ealy $A_E$. Recognition of the $A_L$ at nt 7156 leads to the cleavage of viral late transcripts L1 and L2 at nt 7175 for polyadenylation (13), but this polyadenylation is restricted in those keratinocytes undergoing terminal differentiation. Since all viral transcripts including viral early and late transcripts are initiated from a promoter positioned upstream of the major early poly (A) signal $A_E$, the viral late transcripts L1 and L2 must bypass the early $A_E$ signal to preferentially use the downstream $A_L$ signal for their polyadenylation. Keratinocyte differentiation probably increases the readthrough of the early $A_E$ signal for late gene expression. This might be in part due to changes in activities and levels of polyadenylation and other RNA processing factors, in addition to viral DNA amplification and induction of the late promoter $P_L$. Thus, it is conceivable that efficient

usage of the viral early $A_E$ signal would regulate the expression of viral late genes. However, mutation of the $A_E$ has no effect on the levels of viral late transcripts and polyadenylation at the $A_L$, rather it induces preferential use of a noncanonical polyadenylation site UAUAUA, approximately 100 nts upstream of the major $A_E$ (57). It has been noted that all viral early transcripts are indeed polyadenylated mainly at the viral early $A_E$. By doing so, the viral early transcripts terminate their transcription within the late region beginning approximately 1000 nts downstream of the $A_E$ before reaching the $A_L$ (58) and undergo an endonucleolytic cleavage at the nt 4203 with a little site heterogeneity for polyadenylation.

Viral cis-elements also control polyadenylation of papillomavirus late transcripts (59-61). This was initially discovered in BPV-1 late transcripts. A 53-nt negative regulatory element identified in the BPV-1 late 3' UTR upstream of the $A_L$ (Figure 1) contains a sequence motif homologous to a 5' ss or U1 binding site and inhibits poly (A) tail addition, leading to destabilization of viral late transcripts (59,62). The 53-nt element does not appear to function by destabilizing polyadenylated cytoplasmic RNA (59). Similar U1 binding sites also exist in the HPV16 and HPV31 late 3' UTR and inhibit polyadenylation of those late transcripts (61,62). The mechanism of inhibition involves U1 snRNP binding. In the presence of high levels of free U1 snRNP, BPV-1 late transcripts can be cleaved, but can not be polyadenylated at the RNA 3'-end, and consequently can be degraded rapidly. This mechanism is different from HIV-1, in which the pre-mRNA cleavage is inhibited by the interaction between U1 snRNP and a true 5' ss downstream of the pA site in 5' LTR (63). The inhibition in BPV-1 late transcripts is mediated by the direct interaction between U1 70K protein, one component of U1 snRNP, and poly (A) polymerase (PAP). U1 70K has four PAP inhibitory motifs. Mutation of all four motifs in U1 70K could inhibit PAP to polyadenylate BPV-1 late transcripts (64). However, it remains unknown how the late transcripts can be polyadenylated at the $A_L$ in the presence of 53-nt inhibitory element in differentiated keratinocytes. A logical presumption is that the polyadenylation inhibition of BPV-1 late transcripts may be relieved by the decreased level of free U1 snRNP or its associated factors during the differentiation of keratinocytes (64). U1A is another component of U1 snRNP and is involved in the inhibition of its own pre-mRNA polyadenylation (65) and usage of a secretory pA site in immunoglobulin M (IgM) RNA in B cells. Reduction of U1A during the differentiation of B cells activates the selection of the secretory pA site, suggesting that the repression of polyadenylation by U1 snRNP appears to be associated with cell differentiation (66).

## 5. COUPLING OF RNA SPLICING AND POLYADENYLATION TO RNA EXPORT

After splicing and polyadenylation, a mature mRNA in the nucleus is ready to be exported to the cytoplasm for protein translation. Since RNA export requires association with cellular export factors, identifying such factors essential for BPV-1 early and late mRNA
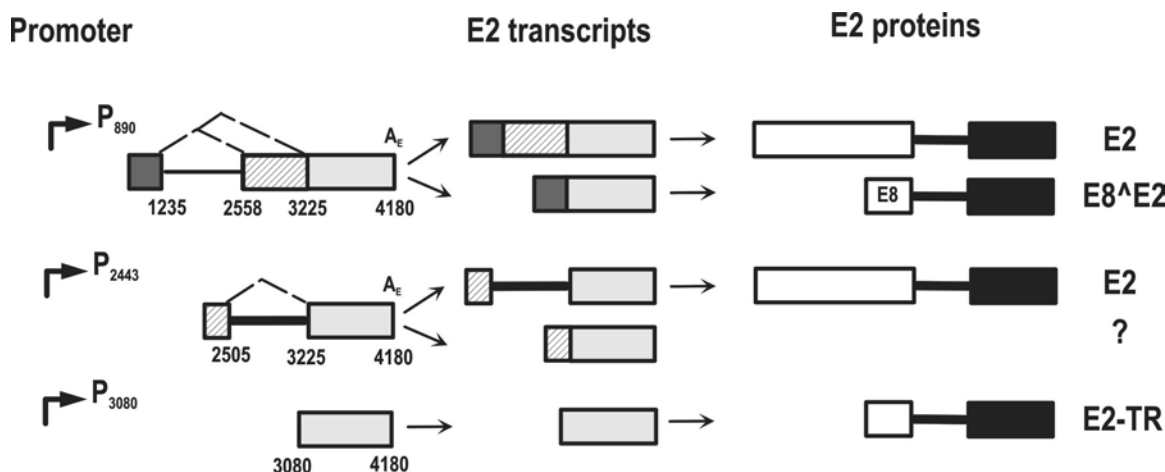
**Figure 4.** Regulation of E2 expression by alternative transcription initiation and alternative splicing. A full-length E2 ORF encodes a transactivator containing a N-terminal transactivation domain (open box), a hinge region (black line), and a C-terminal DNA-binding domain (black box). A full-length E2 protein is encoded either by a $P_{890}$ transcript spliced at nt 2558 3' splice site or by an unspliced $P_{2443}$ transcript. The usage of nt 3225 3' splice site in the transcripts will disrupt the E2 ORF. E8^E2 fusion is encoded by a $P_{890}$ transcript by splicing from nt 1235 5' splice site to nt 3225 3' splice site. E2-TR is encoded by a $P_{3080}$ transcript and contains 162 aa residues from the C-terminal region of the full-length E2. ? indicates a possible coding potential for E5.

export may elucidate other mechanisms of regulation. Various studies indicate that RNA splicing, polyadenylation, and export are coupled events (67,68). RNA splicing creates an exon-exon junction and recruits several proteins including hUpf3B, Aly/REF, SRm160, NXF1, and RNPS1, Magoh, and Y14/RBM8 to bind to an RNA region approximately 20 nts upstream of the junction (69,70). Aly/REF is one of the proteins in the protein complex or exon-junction complex (EJC) and was once thought to be an important factor for all RNA export. Recent studies indicate that Aly/REF plays no role in RNA export (71-74) despite the fact that Aly/REF interacts with several important export factors, including UAP56, a splicing factor and an important mediator in RNA export (75) and NXF1 /TAP, a required RNA transport factor that acts as a heterodimer with p15/Nxt to catalyze translocation of the mature mRNA into the cytoplasm (76). In addition to the roles of the proteins in the EJC, two small members of SR protein family, SRp20 and 9G8, may have some roles in BPV-1 RNA export. SRp20 and 9G8 interact with TAP and mediate RNA export (77-80). Recent studies demonstrated that SRp20 preferentially binds to an A/C-rich SE4 both in viral early and late transcripts in regulation of viral RNA splicing (Jia R and Zheng ZM, unpublished data), but its possible role in viral RNA export can not be excluded.

Several factors involved in RNA polyadenylation may also contribute to viral RNA export. Specifically, PABP binds to the nascent poly (A) tail and facilitates RNA export and translation (55,56). We have noticed that BPV-1 transcription from a BPV-1 late minigene (deletion of a large region with all viral early promoters) in 293 cells terminates gradually from nt 6509 to nt 6840, but only those transcripts that are cleaved and polyadenylated at the $A_E$ site are stable and can be exported into the cytoplasm (Liu XF and Zheng ZM, unpublished data), indicating the importance of the mRNA 3' end polyadenylation in stabilization and export of BPV-1 mRNAs.

## 6. REMARKS AND PERSPECTIVES

BPV-1 has served as a prototype for studying viral life cycle, molecular biology, and pathogenesis of papillomaviruses. This is mainly attributable to the following reasons: (1) The BPV-1 genome had to be one of the first two papillomavirus genomes completely sequenced in the early 1980s (1,81); (2) A simple focus assay was established to study BPV-1 transformation in the late 1970s (82) which allowed researchers to study the viral gene functions involved in BPV-1-induced proliferation and transformation of mouse C127 cells or NIH 3T3 cells in *in vitro* cell culture; (3) A complete transcription map of BPV-1 in mouse C127 cells and bovine fibropapillomas was constructed in the late 1980s (83) and has been the best transcription map for understanding the complex gene expression of other papillomaviruses. The transcription mapping of BPV-1 and later other papillomaviruses has provided a foundation to study viral gene structure and expression, leading to the conclusion that all viral transcripts expressed along with keratinocyte differentiation are bicistronic or polycistronic transcripts, with two or more ORFs in each transcript using either an early or a late poly (A) signal for polyadenylation. This fundamental consensus further led to the identification of viral RNA cis-elements and cellular factors involved in regulation of BPV-1 RNA processing in the early 1990s.

As a prototypical virus, studies of BPV-1 have greatly advanced our knowledge of papillomavirus biology and pathogenesis. Despite the recent shift in focus to more applicable prevention of human papillomavirus infection in the past few years, there are many fundamental questions left in papillomavirus basic research as we discussed in our
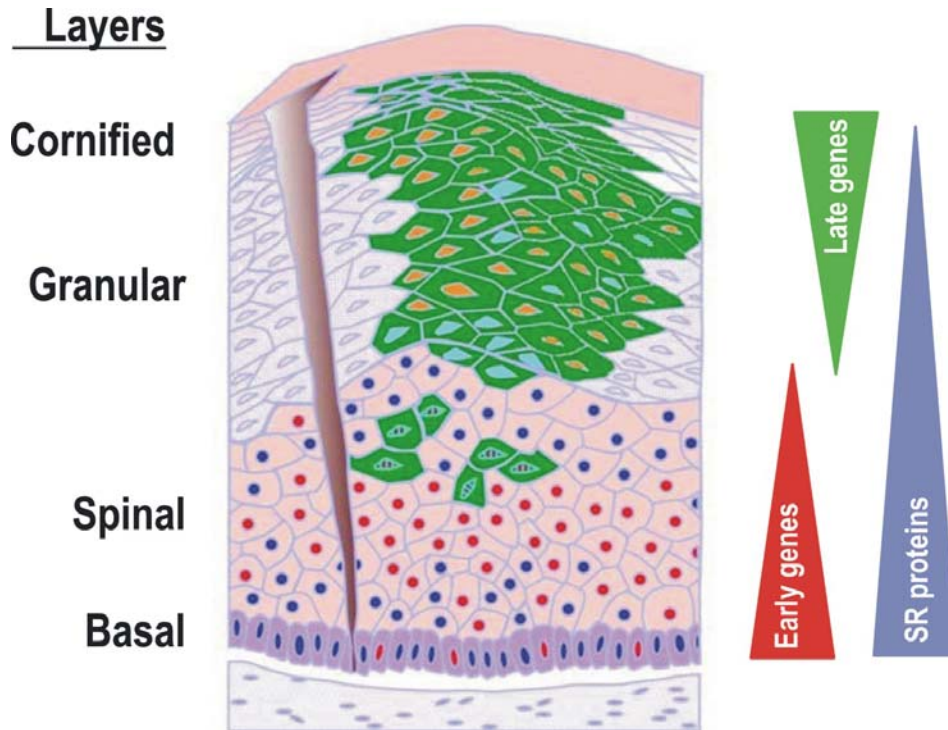
**Figure 5.** SR proteins in BPV-1 early-to-late switch. BPV-1 infects basal keratinocytes via a microtrauma and initiates the expression of viral early genes (red dots) in the undifferentiated basal keratinocytes which contain abundant amount of SR proteins and other cellular splicing factors. After viral DNA replication, late promoter $P_L$ activation and the expression of viral late genes are initiated in highly differentiated keratinocytes in granular and cornified layers (green cells) where the levels of SR proteins and other splicing factors are greatly reduced. Our recent studies show that high level of SRp20, a smallest member of SR protein family, promotes the expression of viral early genes, but suppresses the expression of viral late genes by blocking selection of a BPV-1 late-specific splice site (Jia R and Zheng ZM, unpublished data). Figure is modified with permissions from references (2,84).

previous review (2). With respect to viral gene expression and RNA processing, we have only a limited knowledge on which viral transcript encodes what viral protein in the context of the virus genome, as almost all viral transcripts are transcribed as bicistronic or polycistronic RNAs with two or more ORFs. BPV-1 E6 and E1 ORFs contain a classical U2-type GU-AG intron and L2 ORF a U2-type GC-AG intron. Additional studies are needed to determine what makes intron retention during RNA splicing of these three transcripts or how the transcripts with a retained intron are exported from the nucleus to the cytoplasm for translation into a protein. Transcription of viral late genes and the processing of their transcripts are tightly coupled to keratinocyte differentiation. It will be important to determine what cellular and viral factors contribute to the differentiation-specific regulation of viral late gene expression, which leads to an early-to-late switch of the virus life cycle. The finding that cellular SRp20 partially controls the viral early-to-late switch (Jia R and Zheng ZM, unpublished data) has laid some groundwork for more mechanistic studies. Together with the finding that ASF/SF2 binds the HPV16 3' UTR control element and its expression is regulated during differentiation of virus-infected epithelial cells (51), we propose that a reduced expression of SR proteins or other factors for RNA processing along with keratinocyte differentiation is involved in viral early-to-late switch (Figure 5).

## 7. ACKNOWLEDGEMENTS

## 8. REFERENCES

1. E. Y. Chen, P. M. Howley, A. D. Levinson & P. H. Seeburg, The primary structure and genetic organization of the bovine papillomavirus type 1 genome, Nature 299, 529-534 (1982)

2. Z. M. Zheng & C. C. Baker, Papillomavirus genome structure, expression, and post-transcriptional regulation, Front *Biosci.* 11, 2286-2302 (2006)

3. P. F. Lambert, B. A. Spalholz & P. M. Howley, A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator, *Cell* 50, 69-78 (1987)

4. N. L. Hubbert, J. T. Schiller, D. R. Lowy & E. J.

Androphy, Bovine papilloma virus-transformed cells contain multiple E2 proteins, *Proc.Natl.Acad.Sci.U.S.A* 85, 5864-5868 (1988)

5. J. Choe, P. Vaillancourt, A. Stenlund & M. Botchan, Bovine papillomavirus type 1 encodes two forms of a transcriptional repressor: structural and functional analysis of new viral cDNAs, *J.Virol.* 63, 1743-1755 (1989)

6. W. Waldeck, F. Rosl & H. Zentgraf, Origin of replication in episomal bovine papilloma virus type 1 DNA isolated from transformed cells, *EMBO J.* 3, 2173-2178 (1984)

7. M. Lusky & M. R. Botchan, Transient replication of bovine papilloma virus type 1 plasmids: cis and trans requirements, *Proc.Natl.Acad.Sci.U.S.A* 83, 3609-3613 (1986)

8. B. A. Spalholz, Y. C. Yang & P. M. Howley, Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product, *Cell* 42, 183-191 (1985)

9. B. A. Spalholz, P. F. Lambert, C. L. Yee & P. M. Howley, Bovine papillomavirus transcriptional regulation: localization of the E2-responsive elements of the long control region, *J.Virol.* 61, 2128-2137 (1987)

10. T. H. Haugen, T. P. Cripe, G. D. Ginder, M. Karin & L. P. Turek, Trans-activation of an upstream early gene promoter of bovine papilloma virus-1 by a product of the viral E2 gene, *EMBO J.* 6, 145-152 (1987)

11. S. M. Harrison, K. L. Gearing, S. Y. Kim, A. J. Kingsman & S. M. Kingsman, Multiple cis-active elements in the long control region of bovine papillomavirus type 1 (BPV-1), *Nucleic Acids Res.* 15, 10267-10284 (1987)

12. R. Li, J. Knight, G. Bream, A. Stenlund & M. Botchan, Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome, *Genes Dev.* 3, 510-526 (1989)

13. C. C. Baker & P. M. Howley, Differential promoter utilization by the bovine papillomavirus in transformed cells and productively infected wart tissues, *EMBO J.* 6, 1027-1035 (1987)

14. A. Stenlund, J. Zabielski, H. Ahola, J. Moreno-Lopez & U. Pettersson, Messenger RNAs from the transforming region of bovine papilloma virus type I, *J.Mol.Biol.* 182, 541-554 (1985)

15. Y. C. Yang, H. Okayama & P. M. Howley, Bovine papillomavirus contains multiple transforming genes, *Proc.Natl.Acad.Sci.U.S.A* 82, 1030-1034 (1985)

16. U. Linz & C. C. Baker, Promoters of bovine papillomavirus type 1: *in vitro* activity and utilization, *J.Virol.* 62, 2537-2543 (1988)

17. A. B. Sandler, C. C. Baker & B. A. Spalholz, Sp1 is

critical for basal and E2-transactivated transcription from the bovine papillomavirus type 1 P89 promoter, *J.Gen.Virol.* 77, 189-198 (1996)

18. N. Jareborg, A. Alderborn & S. Burnett, Identification and genetic definition of a bovine papillomavirus type 1 E7 protein and absence of a low-copy-number phenotype exhibited by E5, E6, or E7 viral mutants, *J.Virol.* 66, 4957-4965 (1992)

19. J. Bohl, B. Hull & S. B. Vande Pol, Cooperative transformation and coexpression of bovine papillomavirus type 1 E5 and E7 proteins, *J.Virol.* 75, 513-521 (2001)

20. P. L. Hermonat, B. A. Spalholz & P. M. Howley, The bovine papillomavirus P2443 promoter is E2 trans-responsive: evidence for E2 autoregulation, *EMBO J.* 7, 2815-2822 (1988)

21. B. A. Spalholz, S. B. Vande Pol & P. M. Howley, Characterization of the cis elements involved in basal and E2-transactivated expression of the bovine papillomavirus P2443 promoter, *J.Virol.* 65, 743-753 (1991)

22. H. Ahola, A. Stenlund, J. Moreno-Lopez & U. Pettersson, Promoters and processing sites within the transforming region of bovine papillomavirus type 1, *J.Virol.* 61 , 2240-2244 (1987)

23. P. F. Lambert, N. L. Hubbert, P. M. Howley & J. T. Schiller, Genetic assignment of multiple E2 gene products in bovine papillomavirus-transformed cells, *J.Virol.* 63, 3151-3154 (1989)

24. B. A. Spalholz, J. C. Byrne & P. M. Howley, Evidence for cooperativity between E2 binding sites in E2 trans-regulation of bovine papillomavirus type 1, *J.Virol.* 62, 3143-3150 (1988)

25. J. T. Schiller, E. Kleiner, E. J. Androphy, D. R. Lowy & H. Pfister, Identification of bovine papillomavirus E1 mutants with increased transforming and transcriptional activity, *J.Virol.* 63, 1775-1782 (1989)

26. S. K. Barksdale & C. C. Baker, Differentiation-specific expression from the bovine papillomavirus type 1 P2443 and late promoters, *J.Virol.* 67, 5605-5616 (1993)

27. H. Mishoe, J. N. Brady, M. Radonovich & N. P. Salzman, Simian virus 40 guanine-cytosine-rich sequences function as independent transcriptional control elements *in vitro*, *Mol.Cell Biol.* 4, 2911-2920 (1984)

28. J. Brady, M. Radonovich, M. Vodkin, V. Natarajan, M. Thoren, G. Das, J. Janik & N. P. Salzman, Site-specific base substitution and deletion mutations that enhance or suppress transcription of the SV40 major late RNA, *Cell* 31, 625-633 (1982)

29. S. Tang, M. Tao, J. P. McCoy, Jr. & Z. M. Zheng, The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-

positive cervical cancer cell lines via translation reinitiation, *J.Virol.* 80, 4249-4263 (2006)

30. P. Vaillancourt, T. Nottoli, J. Choe & M. R. Botchan, The E2 transactivator of bovine papillomavirus type 1 is expressed from multiple promoters, *J.Virol.* 64, 3927-3937 (1990)

31. S. Burnett, J. Moreno-Lopez & U. Pettersson, Messenger RNAs from the E1 region of bovine papillomavirus type 1 detected in virus-infected bovine cells, *Nucleic Acids Res.* 15, 8607-8620 (1987)

32. S. Sun, L. Thorner, M. Lentz, P. MacPherson & M. Botchan, Identification of a 68-kilodalton nuclear ATP-binding phosphoprotein encoded by bovine papillomavirus type 1, *J.Virol.* 64, 5093-5105 (1990)

33. W. G. Hubert & P. F. Lambert, The 23-kilodalton E1 phosphoprotein of bovine papillomavirus type 1 is nonessential for stable plasmid replication in murine C127 cells, *J.Virol.* 67, 2932-2937 (1993)

34. S. Barksdale & C. C. Baker, Differentiation-specific alternative splicing of bovine papillomavirus late mRNAs, *J.Virol.* 69, 6553-6556 (1995)

35. K. Neary, B. H. Horwitz & D. DiMaio, Mutational analysis of open reading frame E4 of bovine papillomavirus type 1, *J.Virol.* 61, 1248-1252 (1987)

36. D. L. Black, Mechanisms of alternative pre-messenger RNA splicing, *Annu.Rev.Biochem.* 72, 291-336 (2003)

37. S. F. Jamison & M. A. Garcia-Blanco, An ATP-independent U2 small nuclear ribonucleoprotein particle/precursor mRNA complex requires both splice sites and the polypyrimidine tract, *Proc.Natl.Acad.Sci.U.S.A* 89, 5482-5486 (1992)

38. S. Michaud & R. Reed, An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway, *Genes Dev.* 5, 2534-2546 (1991)

39. B. Ruskin, P. D. Zamore & M. R. Green, A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly, *Cell* 52, 207-219 (1988)

40. D. Staknis & R. Reed, Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly, *Mol.Cell Biol.* 14, 2994-3005 (1994)

41. P. D. Zamore & M. R. Green, Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor, *Proc.Natl.Acad.Sci.U.S.A* 86, 9243-9247 (1989)

42. R. Das, Z. Zhou & R. Reed, Functional association of U2 snRNP with the ATP-independent spliceosomal complex E, *Mol.Cell* 5, 779-787 (2000)

43. A. Watakabe, K. Tanaka & Y. Shimura, The role of exon sequences in splice site selection, *Genes Dev.* 7, 407-418 (1993)

44. Z. M. Zheng, Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression, *J.Biomed.Sci.* 11, 278-294 (2004)

45. Z. M. Zheng, P. He & C. C. Baker, Selection of the bovine papillomavirus type 1 nucleotide 3225 3' splice site is regulated through an exonic splicing enhancer and its juxtaposed exonic splicing suppressor, *J.Virol.* 70, 4691-4699 (1996)

46. Z. M. Zheng, P. J. He & C. C. Baker, Structural, functional, and protein binding analyses of bovine papillomavirus type 1 exonic splicing enhancers, *J.Virol.* 71, 9096-9107 (1997)

47. Z. M. Zheng, M. Huynen & C. C. Baker, A pyrimidine-rich exonic splicing suppressor binds multiple RNA splicing factors and inhibits spliceosome assembly, *Proc.Natl.Acad.Sci.U.S.A* 95, 14088-14093 (1998)

48. Z. M. Zheng, P. J. He & C. C. Baker, Function of a bovine papillomavirus type 1 exonic splicing suppressor requires a suboptimal upstream 3' splice site, *J.Virol.* 73, 29-36 (1999)

49. Z. M. Zheng, E. S. Reid & C. C. Baker, Utilization of the bovine papillomavirus type 1 late-stage-specific nucleotide 3605 3' splice site is modulated by a novel exonic bipartite regulator but not by an intronic purine-rich element, *J.Virol.* 74, 10612-10622 (2000)

50. X. Liu, A. Mayeda, M. Tao & Z. M. Zheng, Exonic splicing enhancer-dependent selection of the bovine papillomavirus type 1 nucleotide 3225 3' splice site can be rescued in a cell lacking splicing factor ASF/SF2 through activation of the phosphatidylinositol 3-kinase/Akt pathway, *J.Virol.* 77, 2105-2115 (2003)

51. M. G. McPhillips, T. Veerapraditsin, S. A. Cumming, D. Karali, S. G. Milligan, W. Boner, I. M. Morgan & S. V. Graham, SF2/ASF binds the human papillomavirus type 16 late RNA control element and is regulated during differentiation of virus-infected epithelial cells, *J.Virol.* 78, 10598-10605 (2004)

52. K. Ryan, O. Calvo & J. L. Manley, Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease, *RNA* 10, 565-573 (2004)

53. C. R. Mandel, S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley & L. Tong, Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease, *Nature* 444, 953-956 (2006)

54. G. M. Gilmartin, Eukaryotic mRNA 3' processing: a common means to different ends, *Genes Dev.* 19, 2517-2521 (2005)

55. D. A. Mangus, M. C. Evans & A. Jacobson, Poly (A)-binding proteins: multifunctional scaffolds for the post-

transcriptional control of gene expression, *Genome Biol.* 4, 223 (2003)

56. J. A. Chekanova & D. A. Belostotsky, Evidence that poly (A) binding protein has an evolutionarily conserved function in facilitating mRNA biogenesis and export, *RNA* 9, 1476-1490 (2003)

57. E. M. Andrews & D. DiMaio, Hierarchy of polyadenylation site usage by bovine papillomavirus in transformed mouse cells, *J.Virol.* 67, 7705-7710 (1993)

58. C. C. Baker & J. S. Noe, Transcriptional termination between bovine papillomavirus type 1 (BPV-1) early and late polyadenylation sites blocks late transcription in BPV-1-transformed cells, *J.Virol.* 63, 3529-3534 (1989)

59. P. A. Furth & C. C. Baker, An element in the bovine papillomavirus late 3' untranslated region reduces polyadenylated cytoplasmic RNA levels, *J.Virol.* 65, 5806-5812 (1991)

60. M. D. Koffa, S. V. Graham, Y. Takagaki, J. L. Manley & J. B. Clements, The human papillomavirus type 16 negative regulatory RNA element interacts with three proteins that act at different posttranscriptional levels, *Proc.Natl.Acad.Sci.U.S.A* 97, 4677-4682 (2000)

61. S. A. Cumming, C. E. Repellin, M. McPhillips, J. C. Radford, J. B. Clements & S. V. Graham, The human papillomavirus type 31 late 3' untranslated region contains a complex bipartite negative regulatory element, *J.Virol.* 76, 5993-6003 (2002)

62. P. A. Furth, W. T. Choe, J. H. Rex, J. C. Byrne & C. C. Baker, Sequences homologous to 5' splice sites are required for the inhibitory activity of papillomavirus late 3' untranslated regions, *Mol.Cell Biol.* 14, 5278-5289 (1994)

63. S. Vagner, U. Ruegsegger, S. I. Gunderson, W. Keller & I. W. Mattaj, Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP, *RNA* 6, 178-188 (2000)

64. S. I. Gunderson, M. Polycarpou-Schwarz & I. W. Mattaj, U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly (A) polymerase, *Mol.Cell* 1, 255-264 (1998)

65. W. C. Boelens, E. J. Jansen, W. J. van Venrooij, R. Stripecke, I. W. Mattaj & S. I. Gunderson, The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA, *Cell* 72, 881-892 (1993)

66. J. Ma, S. I. Gunderson & C. Phillips, Non-snRNP U1A levels decrease during mammalian B-cell differentiation and release the IgM secretory poly (A) site from repression, *RNA* 12, 122-132 (2006)

67. N. J. Proudfoot, A. Furger & M. J. Dye, Integrating mRNA processing with transcription, *Cell* 108, 501-512 (2002)

68. R. Reed & H. Cheng, TREX, SR proteins and export of mRNA, *Curr.Opin.Cell Biol.* 17, 269-273 (2005)

69. H. H. Le, D. Gatfield, E. Izaurralde & M. J. Moore, The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay, *EMBO J.* 20, 4987-4997 (2001)

70. M. E. Stroupe, T. O. Tange, D. R. Thomas, M. J. Moore & N. Grigorieff, The three-dimensional arcitecture of the EJC core, *J.Mol.Biol.* 360, 743-749 (2006)

71. D. Gatfield, H. H. Le, C. Schmitt, I. C. Braun, T. Kocher, M. Wilm & E. Izaurralde, The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila, *Curr.Biol.* 11, 1716-1721 (2001)

72. T. H. Jensen, J. Boulay, M. Rosbash & D. Libri, The DECD box putative ATPase Sub2p is an early mRNA export factor, *Curr.Biol.* 11, 1711-1715 (2001)

73. D. Longman, I. L. Johnstone & J. F. Caceres, The Ref/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans, *RNA* 9, 881-891 (2003)

74. M. MacMorris, C. Brocker & T. Blumenthal, UAP56 levels affect viability and mRNA export in Caenorhabditis elegans, *RNA.* 9, 847-857 (2003)

75. M. L. Luo, Z. Zhou, K. Magni, C. Christoforides, J. Rappsilber, M. Mann & R. Reed, Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly, *Nature* 413, 644-647 (2001)

76. I. C. Braun, A. Herold, M. Rode, E. Conti & E. Izaurralde, Overexpression of TAP/p15 heterodimers bypasses nuclear retention and stimulates nuclear mRNA export, *J.Biol.Chem.* 276, 20536-20543 (2001)

77. Y. Huang & J. A. Steitz, Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA, *Mol.Cell* 7, 899-905 (2001)

78. Y. Huang, R. Gattoni, J. Stevenin & J. A. Steitz, SR splicing factors serve as adapter proteins for TAP-dependent mRNA export, *Mol.Cell* 11, 837-843 (2003)

79. Y. Huang, T. A. Yario & J. A. Steitz, A molecular link between SR protein dephosphorylation and mRNA export, *Proc.Natl.Acad.Sci.U.S.A* 101, 9666-9670 (2004)

80. Y. Hargous, G. M. Hautbergue, A. M. Tintaru, L. Skrisovska, A. P. Golovanov, J. Stevenin, L. Y. Lian, S. A. Wilson & F. H. Allain, Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8, *EMBO J.* 25, 5126-5137 (2006)

81. O. Danos, M. Katinka & M. Yaniv, Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae, *EMBO J.* 1, 231-236 (1982)

82. I. Dvoretzky, R. Shober, S. K. Chattopadhyay & D. R. Lowy, A quantitative *in vitro* focus assay for bovine papilloma virus, *Virology* 103, 369-375 (1980)

83. C. C. Baker & C. Calef, Maps of papillomavirus mRNA transcripts, *in* "The human papillomaviruses compendium" (Los Alamos National Laboratory, Ed.), Los Alamos National Laboratory, (1995)

84. K. Middleton, W. Peh, S. Southern, H. Griffin, K. Sotlar, T. Nakahara, A. El Sherif, L. Morris, R. Seth, M. Hibma, D. Jenkins, P. Lambert, N. Coleman & J. Doorbar, Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers, *J.Virol.* 77, 10186-10201 (2003)

**Send correspondence to:** Zhi-Ming Zheng, 10 Center Dr. Rm. 6N106, Bethesda, MD 20892-1868, Tel: 301-594-1382, Fax: 301-480-8250, E-mail: zhengt@exchange.nih.gov